??
??
??
??

??

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

??

??

$$\sigma'(x) = \sigma(x)[1-\sigma(x)]$$

$$S(x) = \frac{e^{x_i}}{\sum_{j=1}^{k} e^{x_i}}$$

$$f(x) = max(0, x)$$

$$if, x < 0, f(x) = 0$$

$$else f(x) = x$$

$$g_t = \delta_\theta f(\theta)$$

$$m_t = \beta(m_{t-1}) + (1-\beta_1)(\nabla w_t)$$

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t}$$

$$v_t = \beta_2(v_t - 1) + (1-\beta_2)(\nabla w_t)^2$$

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t}$$

$$\omega_{t-1} = \omega_t - \frac{\eta}{\sqrt{\hat{v}_t - \epsilon}}\hat{m}_t$$

(u_t)

y_i

$\hat{y_i}$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y}_i)^2$$

??
$_surface-$
$2 Error surface$[**??**]
$\theta\theta$**??**$\Delta\theta\theta$
$\eta\Delta\theta$**??**

$$\theta = \theta + \eta \cdot \Delta\theta$$

$_vector theta vector$[**??**]

$\theta \Delta\theta$
$L_{(\theta+\eta u)} =$
$L_{(\theta)} +$
$\eta u^T \cdot$
$\bigtriangledown_\theta L_{(\theta)} +$
$\frac{\eta^2}{2!} u^T \cdot$
$\bigtriangledown^2 L_{(\theta)} u +$
$\frac{\eta^3}{3!} \ldots +$
$\frac{\eta^4}{4!} \ldots +$
$\frac{\eta^n}{n!} \ldots$
$\theta \eta \eta^2 <<$
1
$L_{(\theta+\eta u)} =$
$L_{(\theta)} +$
$\eta u^T \cdot$
$\bigtriangledown_\theta L_{(\theta)} [\eta istypicallysmall, so \eta^2, \eta^3, \cdots \rightarrow$
$0]$
$L(\theta +$
$\eta u) L(\theta) L(\theta +$
$\eta u) L(\theta) <$
$0 u^T \cdot$
$\bigtriangledown \theta L(\theta) \bigtriangledown \theta L(\theta) \beta$
$\cos(\beta) =$
$\frac{u^T \cdot \bigtriangledown_\theta L_{(\theta)}}{|u^T||\bigtriangledown_\theta L_{(\theta)}|}$
$\cos(\theta)$
$-1 <$
$\cos(\beta) =$
$\frac{u^T \cdot \bigtriangledown_\theta L_{(\theta)}}{|u^T||\bigtriangledown_\theta L_{(\theta)}|} \leq$
1
$k =$
$|u^T||\bigtriangledown\theta$
$L_{(\theta)}|$
$-k \leq$
$k \cos(\beta) =$
$u^T \cdot$
$\bigtriangledown_\theta L_{(\theta)} \leq$
k

$(L(\theta+$
$\eta u)L(\theta) <$
$0)$
$uT\cdot$
$\bigtriangledown\theta L(\theta)\cos(\beta)\beta$
$180°\theta$
$\theta\theta$
$w_{t=1} =$
w_t-
$\eta\bigtriangledown$
w_t
$b_{t=1} =$
b_t-
$\eta\bigtriangledown$
b_t
$whereatw =$
$w_t, b =$
b_t
$\{\bigtriangledown w_t =$
$\frac{\partial L_{(\theta)}}{\partial w}\bigtriangledown$
$b_t =$
$\frac{\partial L_{(\theta)}}{\partial b}$
??
θ
$\theta =$
$\theta-$
$\eta\cdot$
$\bigtriangledown_\theta L_{(\theta)}$
$\theta\theta$
η
$_epochs) :$
$params_grad =$
$evaluate_gradient(loss_function, data, params)params =$
$paramslearning_rate*$
$params_grad$

$\theta\eta$
??
$J(\theta, x^i, y^i)\theta$
$J(w, b, \theta, x^i, y^i)$
$\theta =$
$\theta -$
$\eta \cdot$
$\bigtriangledown_\theta J_{(\theta, x^i, y^i)}$
??
$_fluctuation SGD fluctuation$[**??**]
$_epochs):$
$np.random.shuffle(data) for example in data:$
$params_grad =$
$evaluate_gradient(loss_function, example, params) params =$
$params learning_rate *$
$params_grad$
??
$\theta =$
$\theta -$
$\eta \cdot$
$\bigtriangledown_\theta J_{(\theta, x^{(i:i+n)}, y^{(i:i+n)})}$
$_epochs):$
$np.random.shuffle(data) for batch in get_batches(data, batch_size =$
$50):$
$params_grad =$
$evaluate_gradient(loss_function, batch, params) params =$
$params learning_rate params_grad$
??
??
??
??

γ
γ
$v_t =$
$\gamma v_{t-1} +$
$\eta \cdot$
$\bigtriangledown_\theta J_{(\theta)}$
$\theta =$
$\theta -$
v_t
$\gamma <$
1
$\gamma v_{t-1} \theta -$
$\gamma v_{t-1} theta$
$v_t =$
$\gamma v_{t-1} +$
$\eta \cdot$
$\bigtriangledown_\theta J_{(\theta - \gamma v_{t-1})}$
$\theta =$
$\theta -$
v_t
??
γ**??**
??
$theta theta \eta$
$theta$
η
g_t
$theta$
$g_{t,i} =$
$\bigtriangledown_\theta J_{(\theta_{t,i})}$
θ_i
$\theta_{t+1,i} =$
$\theta_{t,i} -$
$\eta \cdot$
$g_{t,i}$
$\theta_i \theta_i \eta$
$\theta_{t+1,i} =$
$\theta_{t,i} -$
$\frac{\eta}{\sqrt{G_{t,ii} + \epsilon}}$
$G_t \in$
$R^{d \times d}$
$theta$
$\epsilon \epsilon 10^{-8} G_t theta G_t g_t$
$\theta_{t+1} =$
$\theta_t -$
$\frac{\eta}{\sqrt{G_t + \epsilon}} \cdot$
g_t
??
$E[g^2]_t$
γ
$E[g^2]_t =$
$\gamma E[g^2]_{t-1} +$
$(1-$
$\gamma) g_t^2$
γ
$\triangle \theta_t$
$\triangle \theta_t =$

??

$_{F}low_{o}f_{R}einforcement_{L}earning$

$_entire_interaction_process$
??
$\gamma\lambda$
$\gamma\gamma\gamma\gamma\gamma$
λ
λ

$R(s_t = s) = E[r_t|s_t = s]$

$\gamma \in [0,1]$

$G_t = R_{t+1}+\gamma R_{t+2}+\gamma^2 R_{t+3}+\gamma^3 R_{t+4}+...+\gamma^{T-t-1} R_T$

$V_t(S) = E[G_t|s_t = s]$

$P(s_{s+1} = s'|s_t = s, a_t = a)$

γ
$\gamma =$
$1 \gamma =$
0

$$V(s) = R(s)+\gamma \sum_{s' \in S} P(s'|s)V(s')$$

$R(s)$
$\gamma \sum_{s' \in S} P(s'|s)V(s')$
$V = (1-\gamma P)^{-1} R$

$O(N^3)$

$$g = \sum_{i=t}^{H-1} \gamma^{1-t} r_i$$

$G_t \leftarrow G_t + g, i \leftarrow i+1$

$$V_t(s) \leftarrow \frac{G_t}{N}$$

$P(s_{s+1} =$
$s'|s_t =$
$s, a_t =$
$a)$

$R(s_t = s, a_t = a) = E[r_t|s_t, a_t = a]$

$\gamma \in [0,1]$

$\pi(a|s) = P(a_t = a|s_t = s)$

$P^{\pi}(s's) \overset{\longleftrightarrow}{} \sum_{a\in A}\pi(a|s)P(s'|s,a)$

$P^{\pi}(s) \quad \sum_{a\in A}\pi(a|s)P(s,a)$

$v^{\pi}(s)$

$v^{\pi}(s) = E[G_t|s_t = s]$

$= E[R_{t+1}+\gamma v^{\pi}(s_{t+1})|s_t = s]$

$$= \sum_{a\in A}\pi(a|s)q^{\pi}(s,a)$$

${}_t{}^{o}{}_s{}^{\pi}$

$$v^{\pi}(s) = \sum_{a\in A}\pi(a|s)(R(s,a)+\gamma\sum_{s'\in s}P(s'|s,a)v^{\pi}(s'))$$

$q^{\pi}(s)$

$v^{\pi}(s) = E[G_t|s_t = s]$

$= E[R_{t+1}+\gamma v^{\pi}(s_{t+1})|s_t = s]$

$= \sum_{a \in A} \pi(a|s) q^{\pi}(s, a)$

$_{p}ifunction^{\pi}$
$q^{\pi}(s, a) = R(s, a)+\gamma \sum_{s' \in S} P(s'|s, a) \sum_{a' \in A} \pi(a'|s') q^{\pi}(s', a')$

π
S_t, A_t, R_t
γ
G_t
$G_t =$
$\sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$
$P(s', r|s, a)$
$\pi(a|s)$
$\pi_\theta(.)\theta$
$\mu(s)\pi s$
$V(s)$
$V^\pi(s)V^\pi(s) =$
$E_{a\sim\pi}[G_t|S_t =$
$s]$
$Q(s, a)$
$Q_w(.)$
$Q^\pi(s, a)Q^\pi(s, a) =$
$E_{a\sim\pi}[G_t|S_t =$
$s, A_t =$
$a]$
$A(s, a)$
$A(s, a) =$
$Q(s, a)-$
$V(s)$
θ

$$J(\theta) = \sum_{s\in\mathcal{S}} d^\pi(s)V^\pi(s) = \sum_{s\in\mathcal{S}} d^\pi(s) \sum_{a\in\mathcal{A}} \pi_\theta(a|s)Q^\pi(s, a)$$

$\nabla_\theta J(\theta)$
$d^\pi(.)$

$$\nabla_\theta J(\theta) = \nabla_\theta \sum_{s\in\mathcal{S}} d^\pi(s) \sum_{a\in\mathcal{A}} Q^\pi(s, a)\pi_\theta(a|s)$$

$$\propto \sum_{s\in\mathcal{S}} d^\pi(s) \sum_{a\in\mathcal{A}} Q^\pi(s, a)\nabla_\theta\pi_\theta(a|s)$$

$$\nabla_\theta J(\theta) = \nabla_\theta V^\pi(s_0) = \sum_{s\in\mathcal{S}} d^\pi(s) \sum_{a\in\mathcal{A}} \pi_\theta(a|s)Q^\pi(s, a)\frac{\nabla_\theta\pi_\theta(a|s)}{\pi_\theta(a|s)}$$

$$= E_\pi[Q^\pi(s, a)\nabla_\theta \ln \pi_\theta(a|s)]; Because(\ln x)' = 1/x$$

$*$
$\pi_\theta E_\pi E_{s\sim d\pi, a\sim\pi\theta}$
$\nabla_\theta J(\theta) =$
$\nabla_\theta V^\pi(s_0)$
$\nabla_\theta V^\pi(s)$
$=$
$\nabla_\theta(\sum_{a\in\mathcal{A}} \pi_\theta(a|s)Q^\pi(s, a))$
$=$
$\sum_{a\in\mathcal{A}} \Big(\nabla_\theta\pi_\theta(a|s)Q^\pi(s, a)+$
$\pi_\theta(a|s)\nabla_\theta Q^\pi(s, a)\Big);$
$=$
$\sum_{a\in\mathcal{A}} \Big(\nabla_\theta\pi_\theta(a|s)Q^\pi(s, a)+$
$\pi_\theta(a|s)\sum_{s'} P(s'|s, a)\nabla_\theta V^\pi(s')\Big)_{;P(s'|s,a)=\sum_r P(s',r|s,a)}$
$\phi(s) =$
$\sum_{a\in\mathcal{A}} \nabla_\theta\pi_\theta(a|s)Q^\pi(s, a)$
$\rho^\pi(s \to$
$x, k+$
$1) =$
$\sum_{s'} \rho^\pi(s \to$
$s', k)\rho^\pi(s' \to$
$x, 1)$
$\nabla_\theta V^\pi(s)$
$=$
$\phi(s)+$
$\sum_a \pi_\theta(a|s) \sum_{s'} P(s'|s, a)\nabla_\theta V^\pi(s')$
$=$
$\phi(s)+$
$\sum_{s'} \sum_a \pi_\theta(a|s)P(s'|s, a)\nabla_\theta V^\pi(s')$
$=$
$\phi(s)+$
$\sum_{s'} \rho^\pi(s \to$
$s', 1)\nabla_\theta V^\pi(s')$